

Федеральное агентство научных организаций

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ  
УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР  
«КРАСНОЯРСКИЙ НАУЧНЫЙ ЦЕНТР СИБИРСКОГО ОТДЕЛЕНИЯ  
РОССИЙСКОЙ АКАДЕМИИ НАУК»

Институт вычислительного моделирования СО РАН – обособленное  
подразделение ФИЦ КНЦ СО РАН

ГРНТИ 20.53.17

№ (ЦИТИС) АААА-А18-118011890021-4

ИНВ № 0723/2017

УТВЕРЖДАЮ

Директор ФИЦ КНЦ СО РАН

\_\_\_\_\_ Волков Н.В.  
\_\_\_\_\_ 201\_\_ г.

ОТЧЕТ

О ВЫПОЛНЕНИИ ПРОЕКТА

«МЕТОДЫ И ТЕХНОЛОГИИ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ  
И ПОСТРОЕНИЯ ПРОГРАММНО-ТЕХНИЧЕСКИХ КОМПЛЕКСОВ И  
ИНТЕГРИРОВАННЫХ СИСТЕМ»  
(промежуточный)

Номер проекта в ИСГЗ ФАНО 0356-2016-0723

Приоритетное направление: Информационно-телекоммуникационные  
системы

Программа ФНИ (номер и наименование): IV.35 Когнитивные системы и  
технологии, нейроинформатика и биоинформатика, системный анализ,  
искусственный интеллект, системы распознавания образов, принятие  
решений при многих критериях

Протокол Ученого совета \_\_\_\_\_

№ \_\_\_\_\_ от «\_\_» \_\_\_\_\_ 201\_\_ г.

Руководитель проекта

д.т.н., профессор

\_\_\_\_\_ Л.Ф. Ноженкова

"\_\_" \_\_\_\_\_ 2017 г.

Красноярск, 2017

## СПИСОК ИСПОЛНИТЕЛЕЙ

|                                        |       |                                                           |
|----------------------------------------|-------|-----------------------------------------------------------|
| Руководитель темы<br>д.т.н., профессор | <hr/> | Л.Ф. Ноженкова<br>(Введение, Разделы 1,<br>2, Заключение) |
| Исполнители:<br>зам. дир., к.т.н.      | <hr/> | С.В. Исаев<br>(Раздел 3)                                  |
| г.н.с., д.т.н.                         | <hr/> | А.В. Лапко<br>(Раздел 4)                                  |
| в.н.с., д.т.н.                         | <hr/> | В.А. Лапко<br>(Раздел 4)                                  |
| в.н.с., д.ф.-м.н.                      | <hr/> | М.Г. Садовский<br>(Раздел 5)                              |
| с.н.с., к.т.н.                         | <hr/> | А.А. Евсюков<br>(Раздел 1)                                |
| с.н.с., к.т.н.                         | <hr/> | О.С. Исаева<br>(Раздел 2)                                 |
| с.н.с., к.т.н.                         | <hr/> | В.В. Ничепорчук<br>(Раздел 1)                             |
| с.н.с., к.т.н.                         | <hr/> | Т.Г. Пенькова<br>(Раздел 1)                               |
| с.н.с., к.ф.-м.н.                      | <hr/> | М.Ю. Сенашова<br>(Раздел 5)                               |
| н.с., к.т.н.                           | <hr/> | Д.В. Жучков<br>(Раздел 1)                                 |
| н.с., к.т.н.                           | <hr/> | Е.В. Ковязина<br>(Раздел 1)                               |
| н.с., к.т.н.                           | <hr/> | А.В. Коробко<br>(Разделы 1)                               |
| н.с., к.т.н.                           | <hr/> | А.И. Ноженков<br>(Раздел 1)                               |
| н.с.                                   | <hr/> | Д.Д. Кононов<br>(Раздел 3)                                |

|                  |                          |                                |
|------------------|--------------------------|--------------------------------|
| М.Н.С., К.Т.Н.   | _____<br>(подпись, дата) | В.В. Морозов<br>(Раздел 1)     |
| М.Н.С.           | _____<br>(подпись, дата) | Р.В. Вогоровский<br>(Раздел 2) |
| М.Н.С.           | _____<br>(подпись, дата) | А.А. Коробко<br>(Раздел 1)     |
| М.Н.С.           | _____<br>(подпись, дата) | А.А. Марков<br>(Раздел 2)      |
| прогр. 1-ой кат. | _____<br>(подпись, дата) | С.Н. Кочетков<br>(Раздел 1)    |
| прогр. 1-ой кат. | _____<br>(подпись, дата) | И.А. Ларионова<br>(Раздел 3)   |
| прогр. 1-ой кат. | _____<br>(подпись, дата) | А.В. Малышев<br>(Раздел 3)     |
| программист      | _____<br>(подпись, дата) | Н.В. Кулясов<br>(Раздел 3)     |
| инженер          | _____<br>(подпись, дата) | А.А. Сиротинин<br>(Раздел 3)   |
| Нормоконтролер   | _____<br>(подпись, дата) | А.В. Вяткин                    |

## РЕФЕРАТ

Отчёт 41 с., 16 рис., 2 прил.

### ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, АНАЛИТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ, ПРОГРАММНО-ТЕХНИЧЕСКИЕ КОМПЛЕКСЫ, ИНТЕГРИРОВАННЫЕ СИСТЕМЫ.

Объектом исследования являются методы и технологии информационно-аналитической поддержки принятия решений.

Цель работы – создание новых методов аналитической обработки больших объемов данных, развитие и реализация новых технологических подходов к созданию сложных программно-технических комплексов и интегрированных систем разного назначения.

Выполнен комплекс работ по созданию новых и развитию существующих методов и алгоритмов аналитической обработки и интеграции разнородных данных. Проведен анализ вычислительных затрат расчета аналитических показателей при реализации многошаговых методик. Для повышения эффективности алгоритмов расчета предложена метамодель формирования многомерных оценок аналитических показателей на основе спецификации Common Warehouse Metamodel (CWM). Для развития технологии консолидации разнородных данных в структуру централизованного хранилища разработан подход к представлению древовидной структуры XSD файла в виде объектной модели на основе расширения стандарта CWM, выполнена программная реализация предложенного подхода, обеспечивающая графическое описание объектной модели с помощью языка UML. Предложен технологический подход к организации межсистемного информационного взаимодействия между гетерогенными системами, основанный на унифицированном представлении разнородных структур обмена данными, разработана архитектура программного обеспечения межсистемного взаимодействия. Выполнены работы по развитию web-технологий открытых архивов на примере тестовой модели институционального репозитория ИВМ СО РАН. Выполнено развитие оригинальной реализации технологии модельно-ориентированной разработки программного обеспечения (Model Driven Development, MDD), предложена модель взаимодействия клиентской и серверной частей программного комплекса модельно-ориентированной системы, разработан алгоритм формирования пользовательского интерфейса на основе метаданных, описывающих управляющую модель.

Разработан метод автоматизации испытаний на основе имитационной модели командно-программного управления бортовой аппаратуры космического аппарата. Разработаны алгоритмы информационного взаимодействия имитаторов бортовых систем

и алгоритм анализа прохождения команд по контролю телеметрии. На основе построенной имитационной модели разработан метод автоматизации испытаний командно-программного управления бортовой аппаратурой космического аппарата,. Разработанный метод реализован в подсистеме испытаний командно-программного управления бортовой аппаратурой космического аппарата, входящей в состав программного обеспечения контрольно-проверочной аппаратуры. При проведении имитационных экспериментов выполнено исследование особенностей функционирования бортовых систем, разработаны структуры данных для задания команд и телеметрии, сформированы перечень команд и критерии контроля их прохождения, разработаны алгоритмы информационного взаимодействия имитаторов бортовых систем. Разработаны графические интерфейсы программной инфраструктуры имитационного моделирования бортовой аппаратуры спутниковых систем, основанной на стандарте SMP (Simulation Model Portability).

Выполнены исследования по оцениванию различных аспектов кибербезопасности информационно-телекоммуникационных систем. Применены методы кластерного анализа больших данных о доступе пользователей к интернет-сервисам, исследованы способы визуализации полученных результатов. Предложена методика анализа использования ресурсов сети и выявления аномалий, способствующих возникновению угроз кибербезопасности. Исследованы методы интеграции существующих систем информационной безопасности, разработана функциональная схема универсальной сигнатурной системы управления событиями информационной безопасности. Проведено развитие методики разработки защищенных веб-приложений, предложен метод интегральной оценки защищенности веб-приложений на основе векторов групповых оценок критериев международного стандарта Advanced Security Verification Standard.

Разработаны статистические методы распознавания образов и восстановления многомерных зависимостей в условиях больших выборок неоднородных данных, основанные на дискретизации области их определения, принципах коллективного оценивания и гибридного моделирования. Установлены условия асимптотической сходимости соответствующих им решающих функций. На этой основе определены аналитические зависимости показателей эффективности оценивания решающих функций от особенностей исходных данных и используемых методов их обработки. Проведён анализ методов оптимизации непараметрической оценки плотности вероятности Розенблатта-Парзена и дискретизации области значений двумерной случайной величины, направленных на повышение вычислительной эффективности исследуемых моделей и алгоритмов.

Исследованы структуры геномов хлоропластов и цианобактерий. Для каждого фрагмента генома вычислен частотный словарь троек символов без пересечений. Для каждого генома в 64-мерном пространстве частот построены проекции в пространство первых трёх главных компонент. Обнаружено, что подавляющее большинство геномов хлоропластов имеет восьмикластерную структуру. Геномы цианобактерий имеют одинаковую структуру, представляющую собой своеобразный клубок из нитей, образованных точками, соответствующими последовательным фрагментам генома.

Область применения полученных результатов – создание сложных программно-технических комплексов и интегрированных систем для информационно-аналитической поддержки принятия решений в разных прикладных областях.

## СОДЕРЖАНИЕ

|                                                                                                                                                                                                                      |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ВВЕДЕНИЕ.....                                                                                                                                                                                                        | 8  |
| 1 Развитие и апробация технологий создания интегрированных систем в различных областях территориального и корпоративного управления.....                                                                             | 10 |
| 2 Развитие методов автоматизированной поддержки испытаний программно-аппаратных комплексов бортовой аппаратуры космического аппарата.....                                                                            | 18 |
| 3 Разработка методов оценки различных аспектов кибербезопасности информационно-телекоммуникационных систем .....                                                                                                     | 23 |
| 4 Разработка статистических методов распознавания образов и восстановления многомерных зависимостей в условиях неоднородных данных, основанных на принципах коллективного оценивания и гибридного моделирования..... | 27 |
| 5 Выявление новых структурных элементов в символьных последовательностях, соответствующих генетическим данным, на основе изучения взаимного распределения триплетов и октаплетов.....                                | 31 |
| ЗАКЛЮЧЕНИЕ.....                                                                                                                                                                                                      | 35 |
| Приложение А.....                                                                                                                                                                                                    | 38 |
| Приложение Б.....                                                                                                                                                                                                    | 41 |

## ВВЕДЕНИЕ

Развитие технологий построения интегрированных информационно-аналитических систем и программно-технических комплексов опирается на создание новых методов аналитической обработки больших объемов данных. Актуальны исследования, направленные на решение проблемы консолидации данных из гетерогенных источников и их совместной аналитической обработки. В связи с необходимостью разработки сложных комплексных систем остается актуальной проблема развития новых методов интеграции технологий. Значительное количество работ зарубежных авторов и растущий интерес отечественных исследователей подтверждают актуальность проблемы межсистемного информационного взаимодействия. Необходимость совместимости корпоративных систем и их интеграции с внешними ресурсами требуют разработки и развития методов и технологий, обеспечивающих межсистемный обмен данными с возможностью гибкой адаптации к изменениям условий функционирования систем.

Развитие отечественного производства бортовых систем космических аппаратов требует создания интегрированных технологий поддержки производства программно-аппаратных комплексов, включая технологии имитационного моделирования, поддержки программно-аппаратного взаимодействия для проведения испытаний, технологий анализа функционирования бортовых устройств.

Современные информационно-телекоммуникационные системы подвержены растущим рискам, связанным с информационной безопасностью. Актуальность и важность вопросов защиты от сетевых угроз подтверждается исследованиями ведущих антивирусных компаний.

**Цель работы** – создание новых методов аналитической обработки больших объемов данных, развитие и реализация новых технологических подходов к созданию сложных программно-технических комплексов и интегрированных систем разного назначения.

Основные задачи:

1. Развитие и апробация технологий создания интегрированных систем в различных областях территориального и корпоративного управления.
2. Развитие методов автоматизированной поддержки испытаний программно-аппаратных комплексов бортовой аппаратуры космического аппарата.
3. Разработка методов оценки различных аспектов кибербезопасности информационно-телекоммуникационных систем.



4. Разработка статистических методов распознавания образов и восстановления многомерных зависимостей в условиях неоднородных данных, основанных на принципах коллективного оценивания и гибридного моделирования.

5. Выявление новых структурных элементов в символьных последовательностях, соответствующих генетическим данным, на основе изучения взаимного распределения триплетов и октаплетов.

Настоящий отчет представляет результаты за 2017 год по теме «Методы и технологии аналитической обработки данных и построения программно-технических комплексов и интегрированных систем».

*1 Развитие и апробация технологий создания интегрированных систем в различных областях территориального и корпоративного управления.*

Выполнен комплекс работ по созданию новых и развитию существующих методов и алгоритмов аналитической обработки и интеграции разнородных данных.

Проведен анализ вычислительных затрат расчета аналитических показателей при реализации многошаговых методик. На основе формализации базовых операций над OLAP-кубом проанализирована производительность формируемых запросов при различных сочетаниях аналитических операций с точки зрения мощности обрабатываемого куба. Результаты исследования позволили определить оптимальную последовательность операций для повышения эффективности алгоритмов расчета аналитических показателей. На примере данных гидрологического мониторинга определен наиболее эффективный путь извлечения оперативных данных за счет сокращения мощности куба.

Предложена метамодель формирования многомерных оценок показателей природно-техногенной безопасности на основе спецификации Common Warehouse Metamodel (CWM), которая представляет открытый технологический стандарт, обеспечивающий информационное взаимодействие инструментов обработки данных за счет формирования унифицированной структуры метаданных. На основе CWM разработана расширенная метамодель OLAP, определяющая логическую структуру аналитических моделей. Разработана метамодель аналитических операций над OLAP-кубом, позволяющая реализовать алгоритм расчета многомерных оценок показателей. Расширение стандартной метамодели выполнено путем включения дополнительного класса «*Standard*», который задает многомерные нормативные значения показателя, дополнительного атрибута «*isTerritory*» класса «*Dimension*», который обеспечивает привязку данных к измерению территории, а также дополнительной аналитической операции над OLAP-кубом «*Estimation*», которая реализует особенности алгоритма расчета многомерных оценок. В соответствии с расширенной метамоделью, на основе расширяемого языка разметки предложены DTD-описания, определяющие структуру XML-файлов для хранения и передачи аналитических моделей. Предложенные в работе метаописания обеспечивают оперативную актуализацию многомерных моделей и представляют единый формат работы аналитических средств с OLAP-моделями.

С целью обеспечения консолидации данных в централизованном хранилище усовершенствована технология управления хранилищем данных, позволяющая подключать информационные источники для загрузки данных различных форматов,

выбирать регламент актуализации данных, создавать и корректировать процедуры очистки и обновления данных. Предложенная технология обеспечивает двусторонний информационный обмен с корпоративными системами с помощью встроенных механизмов API. Для консолидации разнородных данных в структуру централизованного хранилища разработан подход к представлению древовидной структуры XSD (XML Schema Definition) файла в виде объектной модели на основе расширения стандарта Common Warehouse Metamodel (CWM), определяющего основные классы и ассоциации для XML версии 1.0. В соответствии с классической BUS-архитектурой хранилищ данных информационным ядром базы является «шина» справочников, определяющие аспекты анализа – измерения, которые характеризуются, как правило, строковым типом данных, конечным множеством принимаемых значений и иерархической структурой отношений. Предложены алгоритмы определения идентифицирующих атрибутов, ссылочных полей и восстановления иерархических зависимостей между элементами. Восстановление иерархических зависимостей выполнено путем сопоставления ComplexType, как именованных, так и неименованных и интерпретации структур extension, choice, restriction, pattern и enumeration. Для формализации дополнительных характеристик структурных элементов схемы предложено расширение стандарта CWM. Выполнена программная реализация построения объектной модели по XSD схеме на языке Java в среде разработки IntelliJ IDEA (рисунок 1.1). Построение объектной модели XSD файла обеспечивает основу для выделения аналитических концептов (показателей и измерений) из информационного контекста источника данных, использующего формат XML для хранения и обмена информацией, что позволяет интегрировать данные такого источника в структуру единого централизованного хранилища.

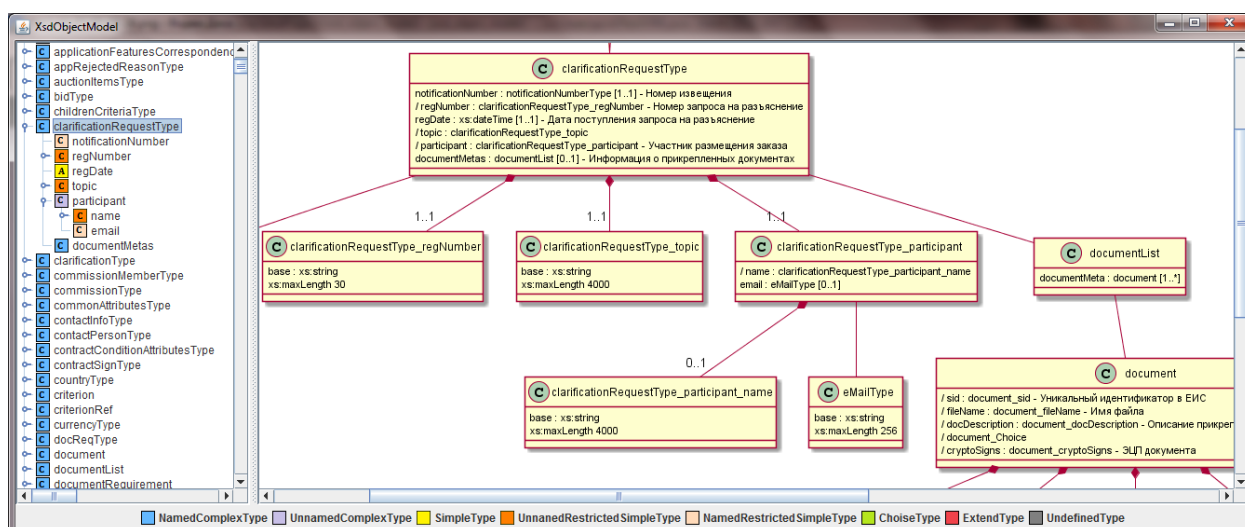


Рисунок 1.1 - Построение объектной аналитической модели для XSD схемы

Предложен технологический подход к организации межсистемного информационного взаимодействия между гетерогенными системами, основанный на унифицированном представлении разнородных структур обмена данными.

Разработана метамодель, определяющая логику и принципы информационного обмена: взаимодействие осуществляется между двумя подсистемами, каждая из которых использует определенный формат обмена данными. Возможность автоматизации сопоставления схем хранения в разных подсистемах обеспечивается унифицированной моделью представления данных, которая задает классы структурных элементов, не зависящие от конкретной реализации (формата, схемы хранения данных), и настраиваемой моделью, которая содержит параметры и условия передачи данных. Разработана структура иерархической унифицированной модели и шаблоны настраиваемой модели для типовых объектов информационного обмена и форматов передачи данных. Разработаны алгоритмы формирования унифицированных представлений и алгоритмы преобразования структур обмена данными в соответствии с настраиваемой моделью. Настраиваемая модель, с точки зрения хранения, представляет собой xml-файл. Предложенный подход обеспечивает автоматизацию двустороннего обмена данными – загрузку в систему сведений из внешних XML-файлов и выгрузку сведений системы в XML-файлы произвольных форматов – без вмешательства в структуру данных и программное обеспечение функционирующей корпоративной системы (рисунок 1.2).

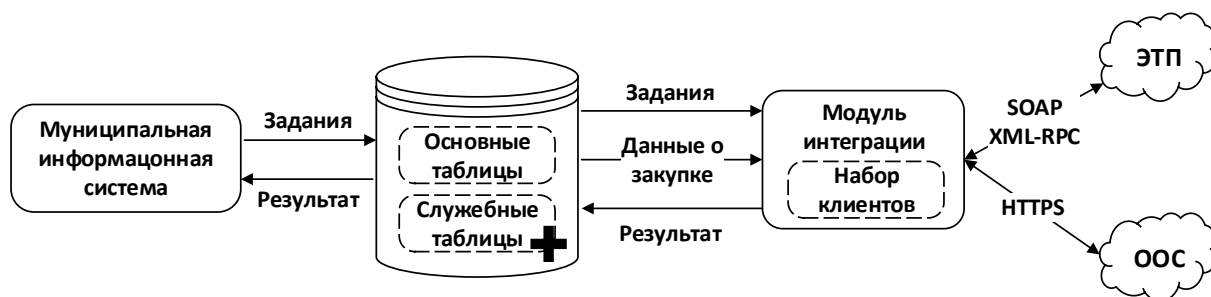


Рисунок 1.2 - Организация межсистемного информационного взаимодействия

В соответствии с предложенным подходом, разработана архитектура программных средств организации межсистемного взаимодействия (рисунок 1.3).

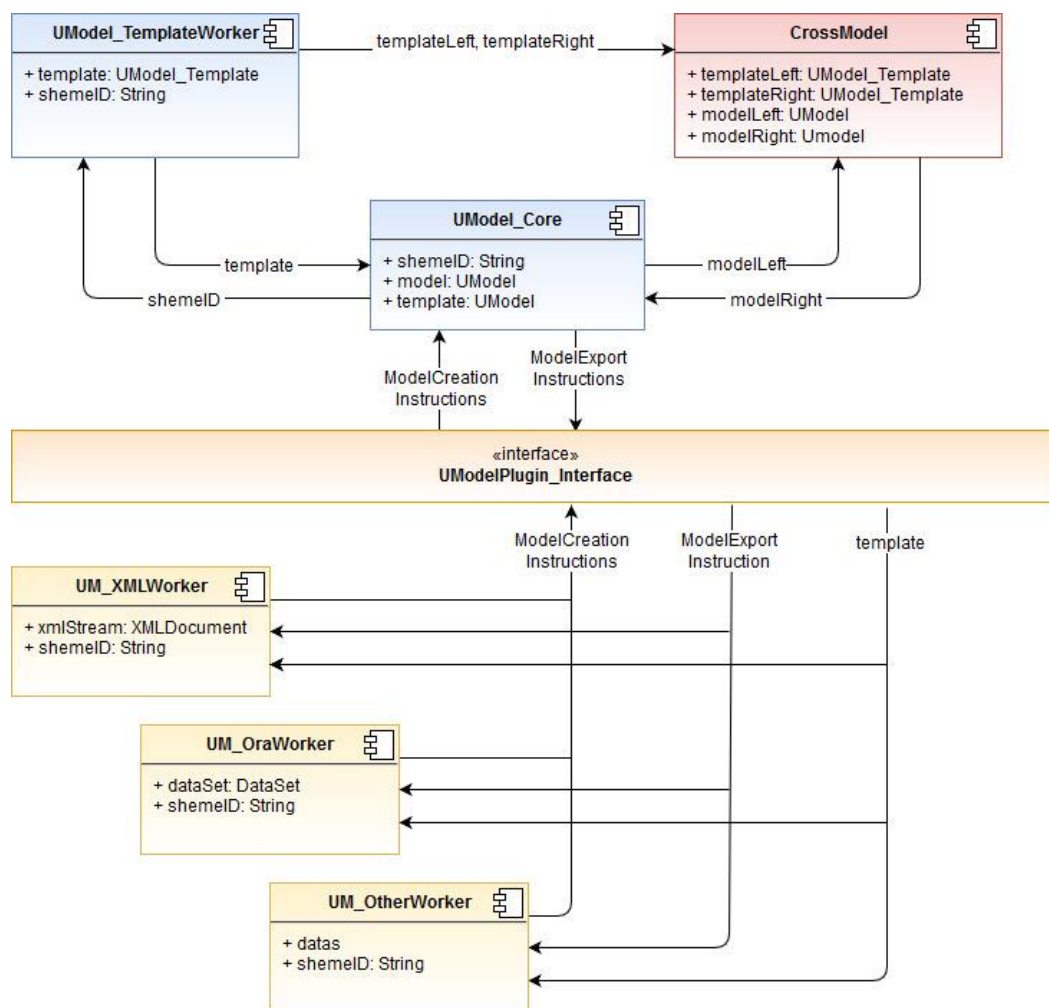


Рисунок.1.3 - Архитектура программных средств межсистемного взаимодействия.

### Диаграмма классов UML

Предложенная архитектура содержит следующие основные компоненты. *UModel\_Core* – компонент, отвечающий за построение унифицированной модели на основе заданной схемы хранения и переданного информационного объекта. В процессе построения создаются экземпляры классов модели различного уровня, устанавливаются связи между экземплярами, выделяются ключевые поля и значения, необходимые для дальнейшей идентификации данных. Каждая модель сопровождается шаблоном, определяющим исходную структуру данных. *UModel\_TemplateWorker* – компонент, отвечающий за работу с шаблонами данных унифицированной модели, который обеспечивает чтение и запись шаблона в хранилище шаблонов, поиск нужного шаблона для потока данных, первичную валидацию потока данных указанному шаблону, а также работу визуальных инструментов для построения шаблона. *UModelPlugin\_Interface* – базовый интерфейс подключаемых модулей, содержащий описание вызова функций, необходимых для построения и наполнения модели из произвольного (при наличии

соответственных плагинов) источника данных. *UModel\_DataPlugin* – подключаемые модули, обеспечивающие подключение к источникам данных, чтение информации и запись данных в требуемый формат. Экземплярами данного модуля являются: *UM\_XMLWorker*, выполняющий обработку данных в формате XML, *UM\_OraWorker*, выполняющий обработку данных в формате реляционной базы (в частности Oracle) и другие модули, зависящие от формата передаваемых данных. *UCrossModel* – компонент, отвечающий за работу с настраиваемой моделью: построение модели, сопоставление унифицированных структур и преобразование одного формата данных в другой на основе анализа шаблонов («левый» описывает схему данных отправителя, «правый» описывает схему данных, получателя). Предложенный технологический подход и архитектура обеспечивает обмен данными между гетерогенными ресурсами в муниципальной системе организации закупок: Автоматизированной системой поддержки размещения муниципального заказа (АСП МЗ), федеральными электронными торговыми площадками и официальным сайтом Единой информационной системы в сфере закупок [zakupki.gov.ru](http://zakupki.gov.ru).

Выполнено развитие web-технологий открытых архивов на примере тестовой модели институционального репозитория ИВМ СО РАН, развернутого на свободно распространяемой платформе DSpace 6.0.

Выполнен импорт описательных метаданных объектов репозитория из системы автоматизации библиотек ИРБИС, в которой осуществляется их первоначальное формирование. Под объектами репозитория понимаются включаемые в него полнотекстовые документы. Произведено сопоставление полей форматов Dublin Core (DC) (включая квалификаторы) и Rusmasrc для обеспечения импорта данных без потерь информативности. Выделены поля метаданных ИРБИС, отсутствующие в базовой схеме DC, и оценена необходимость их наличия в описательных метаданных репозитория. В частности, это касается полей, предназначенных для наукометрической и библиометрической оценок публикаций: динамически обновляемого из внешних источников числа цитирований, импакт-фактора и квартиля источника публикации, списка библиографии и т.п. Произведены тестовый импорт и пополнение списка полей.

На тестовом массиве репозитория выполнена настройка системы поддержки издательского эмбарго для объектов репозитория по широкому спектру условий – от полностью закрытых («темных») объектов, включающих закрытость их описаний, до условно и полностью открытых документов. Технология позволяет разместить научных публикации в открытом доступе на легитимной основе с соблюдением норм российского законодательства об авторском и смежном правах, а также рассмотреть условия свободного лицензирования объектов репозитория на основе Creative Common.

Проведены исследования форм организации связей объектов репозитория в трех направлениях – внутри одной коллекции, между коллекциями одного сообщества и с внешними источниками данных. Задача актуальна, например, при дублировании идентичных объектов в разных коллекциях, а также при организации контролируемых словарей и авторитетных данных. Решение задачи позволит минимизировать дублирование данных и ошибки в полях метаданных объектов.

На тестовом массиве проведена интеграция репозитория в поисковую среду Google Analytic за счет встроенных в базовую программную платформу средств интеграции и механизмов связи с внешними источниками с помощью API. Успешная реализация интеграции репозитория с аналитическими инструментами Google дает надежду на интеграцию в дальнейшем с иными поисковыми системами Интернет, например, Yandex, что несомненно послужит увеличению гибкости системы как в области аналитики, так и в использовании широкого спектра внешних API-сервисов.

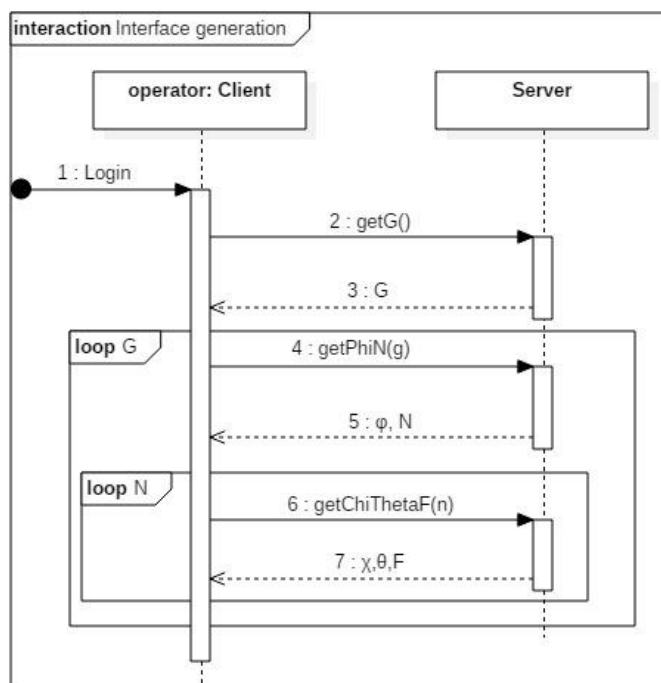
Результаты исследований показали, что институциональные репозитории, формируемые в технологиях архивов открытого доступа, позволяют произвести качественный учет и продвижение научных публикаций, обеспечить открытость результатов научных исследований для мирового сообщества, повысить индивидуальные показатели цитирования. Открытый архив позволит в перспективе успешно развивать связывание данных, а также интегрировать данные репозитория в корпоративные, региональные или национальные проекты открытых архивов, в соответствие с общемировыми тенденциями развития web-технологий. Импорт/экспорт описательных метаданных научных публикаций из/в имеющиеся в организации или внешние библиографические базы данных позволит эффективно решать проблему содержательного наполнения репозитория. Развитие и внедрение web-технологий открытых архивов позволит научно-исследовательским организациям создавать и продвигать собственные ресурсы и результаты исследований, эффективно и согласованно влияя на формирование информационной среды Интернет. Проведенные исследования нашли отражение в разработанных Центральной научной библиотекой ФИЦ КНЦ СО РАН организационных документах «Положение об электронной библиотеке» и «Стратегия развития библиотеки ФИЦ КНЦ СО РАН», определяющих перспективы и технологии формирования информационных ресурсов и систем федеративного поиска.

Выполнено развитие оригинальной реализации технологии модельно-ориентированной разработки программного обеспечения (Model Driven Development, MDD), предложена модель взаимодействия клиентской и серверной частей программного

комплекса модельно-ориентированной системы, разработан алгоритм формирования пользовательского интерфейса на основе метаданных, описывающих управляющую модель.

Построение модельно-ориентированных систем сбора данных в рамках оригинальной реализации подхода к модельно-ориентированной разработке (MDD) включает этапы проектирования моделей абстрактного уровня и моделей прикладного уровня. В отличие от классического подхода предложено объединить процессы формирования управляющей и прикладной моделей и добавить процесс формирования модельно-ориентированной системы, который позволяет получить в результате выполнения всех процессов не программный код, а рабочую систему.

Мета-мета-модель содержит класс объектов  $N$ , класс атрибутов  $F$  и класс групп объектов  $G$ . В рамках мета-мета-модели задан ряд отношений, определяющих связи между экземплярами классов модели. Между объектами определены два вида отношений, вложенность  $\varphi$  и зависимость  $\chi$ . Между объектами и атрибутами определено отношение обладания  $\theta$ . Между объектами и группами – группировка  $\psi$ . Модель взаимодействия веб-клиента и сервера при формировании интерфейса оператора модельно-ориентированной системы приведена на рисунке 1.4.



Рисуноу 1.4 - Модель взаимодействия клиентской и серверной частей  
модельно-ориентированной системы



На первом этапе веб-клиент запрашивает список доступных групп  $getG()$ , в ответ получает  $G$  – множество групп классов объектов. Выбор группы инициирует следующее обращение к серверу  $getPhiN(g)$ . Сервер, основываясь на отношении  $\psi$ , передает веб-клиенту множества объектов  $Ng \subseteq N$ , относящихся к группе  $g \in G$ , и информацию об отношении  $\varphi$ , определяющем внутреннюю иерархию объектов. Пользователь, выбирая конкретный класс объектов  $n \in Ng$ , инициирует обращение веб-клиента к серверу  $getChiThetaF(n)$  для получения списка свойств и связанных объектов для выбранного класса  $n$ . Из диаграммы на рисунке 1.4 видно, что выбор группы и выбор объекта может повторяться несколько раз по желанию пользователя, в результате будет происходить динамическое перестроение системы в соответствии с данными о модели, полученными от сервера.

Как видно из модели взаимодействия клиентской и серверной частей модельно-ориентированной системы, веб-клиент и сервер тесно связаны. Клиент постоянно обращается к серверу для получения данных необходимых для построения и функционирования клиентской системы (включая интерфейс пользователя), сервер формирует метаданные по требованию на основании управляющей модели и отправляет их в формате json. Алгоритм работы веб-клиента при построении интерфейса представлен на рисунке 1.5.

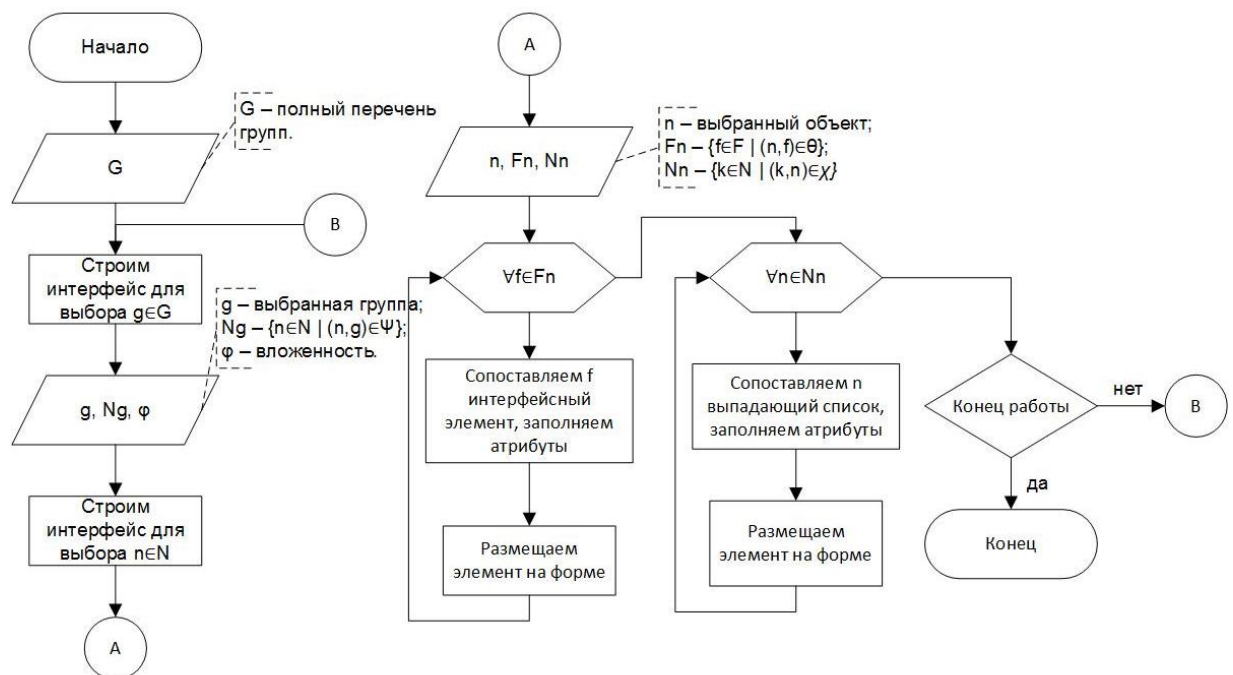


Рисунок 1.5 - Алгоритм построения интерфейса клиентской части модельно-ориентированной системы

Алгоритм начинает работу с получения полного перечня доступных групп  $G$ . Выбор группы  $g \in G$  определяет  $Ng \subseteq N$  – подмножество объектов, которое формируется в соответствии с отношением между объектами и группами  $\psi$ . Кроме подмножества, сервер передает информацию об отношении  $\varphi$ , которое описывает иерархические зависимости между объектами множества  $N$  в целом и  $Ng$  в частности. Согласно алгоритму, после выбора класса объектов  $n$  с сервера запрашивается множество атрибутов  $Fn$ , характерных для объектов этого класса, и множество объектов  $Nn$ , связанных с  $n$  отношением зависимости. В соответствии со спецификацией мета-метамодели, множество атрибутов для выбранного класса объектов определяется на основании отношения  $\theta$ , а множество связанных объектов на основании отношения  $\chi$ . С использованием полученных от сервера данных, веб-клиент генерирует форму для ввода и редактирования данных. Каждому атрибуту  $f$  из множества полученных с сервера атрибутов  $Fn$  ставится в соответствие интерфейсный элемент, заполняются настройки, указанные в управляющей и прикладной моделях. Затем в соответствии с полученными метаданными элемент помещается в интерфейс. Типы интерфейсных элементов определяются на этапе проектирования.

Предложенный алгоритм работы клиентской части системы использован при построении модельно-ориентированной системы с функцией управления метамodelью и возможностью изменения своего функционала силами обычных пользователей. Использование алгоритма позволяет достичь высокого уровня адаптируемости информационной системы. Апробация выполнена в Институте вычислительного моделирования при создании системы учёта результатов научной деятельности и веб-системы сбора данных территориальных центров мониторинга и прогнозирования ЧС.

## *2 Развитие методов автоматизированной поддержки испытаний программно-аппаратных комплексов бортовой аппаратуры космического аппарата.*

Разработан метод автоматизации испытаний на основе имитационной модели командно-программного управления бортовой аппаратуры космического аппарата. В модели построена схема потоков данных, которая позволила определить типы данных, направления информационного взаимодействия и контрольные точки для анализа прохождения и отработки команд (рисунок 2.1).

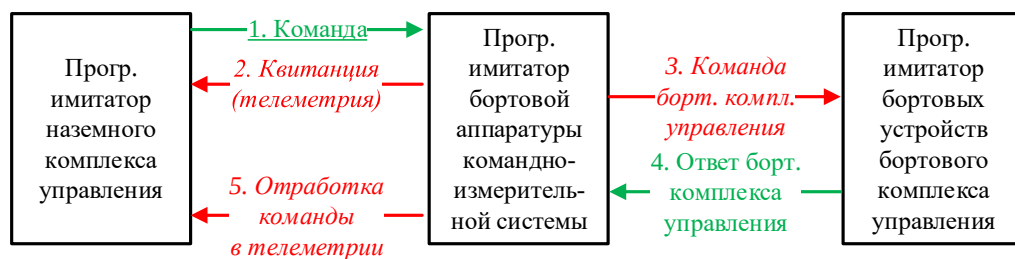


Рисунок 2.1 - Поток данных для имитации командно-программного управления

Разработаны алгоритмы информационного взаимодействия имитаторов бортовых систем (рисунок 2.2) и алгоритм анализа прохождения команд по контролю телеметрии (рисунок 2.3).

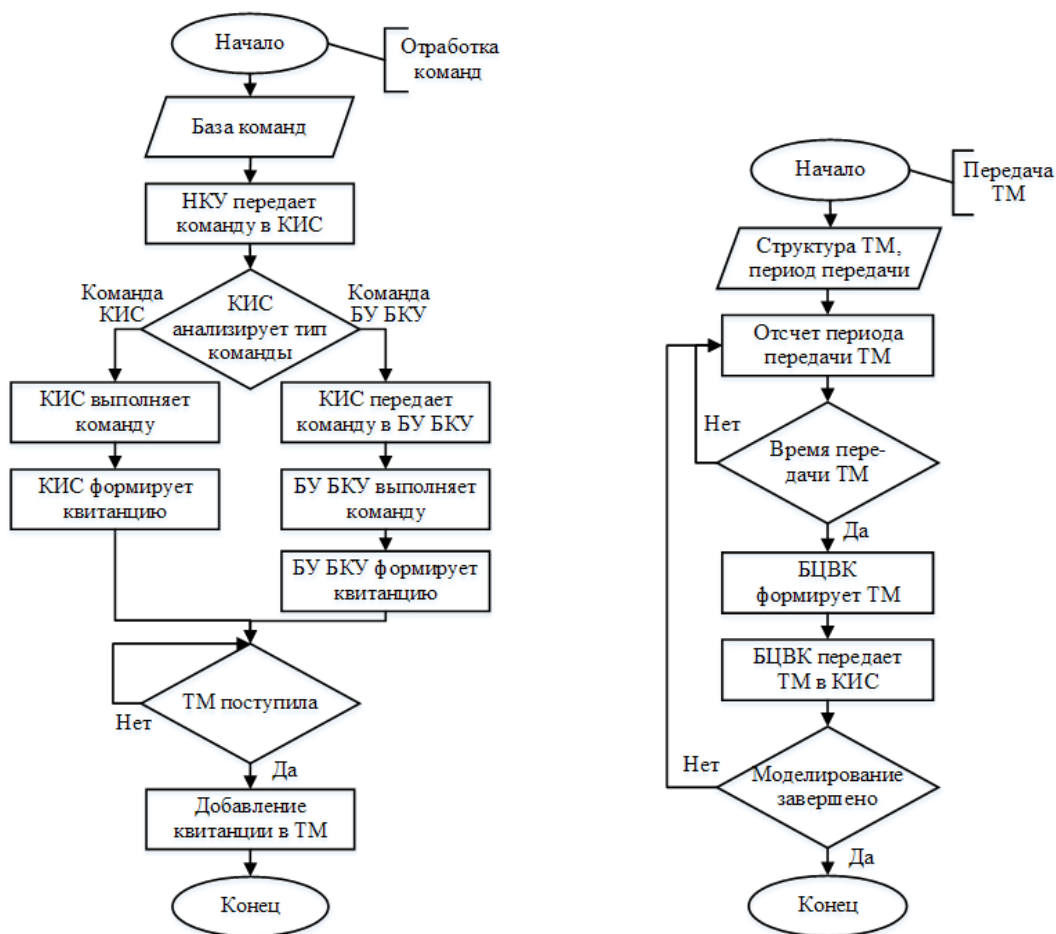


Рисунок 2.2 - Алгоритмы информационного взаимодействия имитаторов бортовых систем

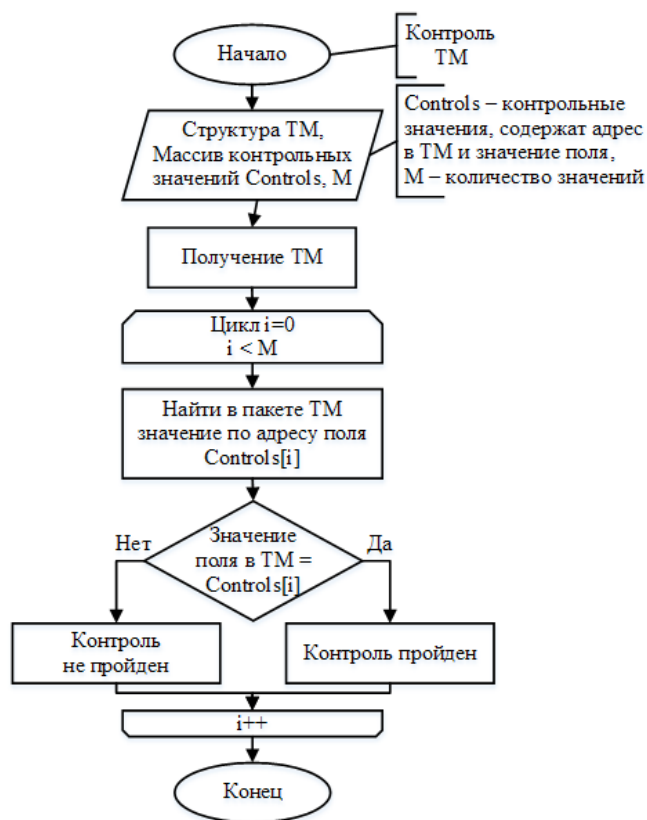


Рисунок 2.3 - Алгоритм анализа прохождения команд по контролю телеметрии

В алгоритмах реализованы функции программных имитаторов: имитатор наземного комплекса управления (НКУ), имитатор командно-измерительной системы (КИС), имитатор бортовых устройств бортового комплекса управления (БУ БКУ) и входящего в его состав бортового цифрового вычислительного комплекса (БЦВК). Информационное взаимодействие бортовых систем включает процесс приема и обработки команд, а также формирования и передачи телеметрии. Реализация алгоритмов информационного взаимодействия бортовых систем позволяет моделировать различные сценарии функционирования командно-измерительной системы, изменять порядок передачи команд. Алгоритм анализа прохождения команд позволяет формировать критерии контроля работы бортовых систем.

На основе построенной имитационной модели разработан метод автоматизации испытаний командно-программного управления бортовой аппаратурой космического аппарата. Испытания выполняются программно-аппаратным комплексом контрольно-проверочной аппаратуры. Новый метод упрощает принцип построения испытательных процедур, позволяя настраивать имитаторы окружения объекта контроля в соответствии с алгоритмом анализа прохождения команд и контрольными точками, выявленными при реализации схемы информационного взаимодействия.

На рисунке 2.4 представлена диаграмма формирования методики проведения испытаний в программном обеспечении контрольно-проверочной аппаратуры. Методики испытаний задаются в виде сценариев, которые определяют последовательности заданий, состоящих из действий. Такая схема задания испытаний бортовой аппаратуры универсальна, но сложна для реализации и использования.

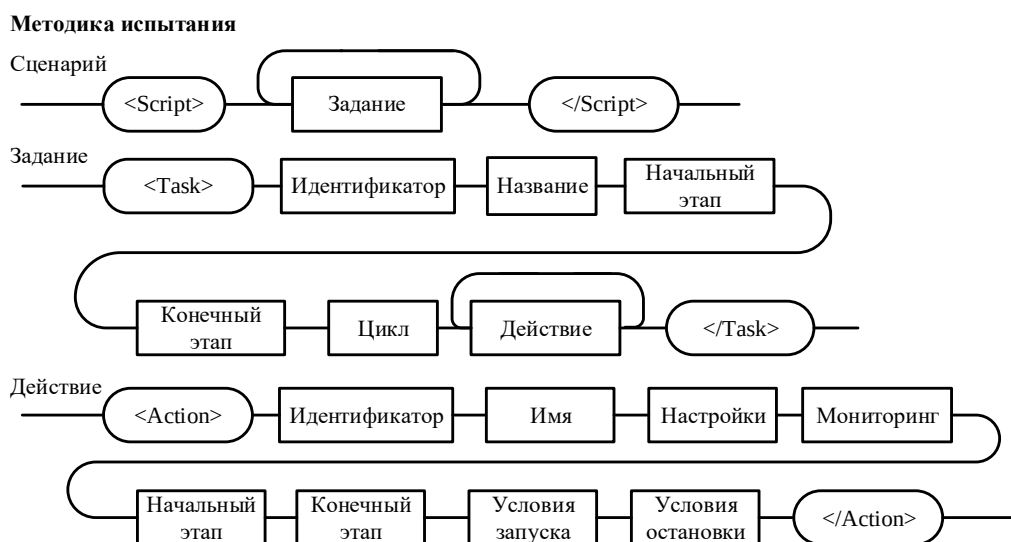


Рисунок. 2.4 - Диаграмма формирования методики испытаний

Анализ имитационной модели позволил упростить схему создания испытательных процедур и выделить контрольные точки информационного взаимодействия. Новый метод автоматизации испытаний расширяет функции программно-аппаратного комплекса контрольно-проверочной аппаратуры, упрощает процесс формирования методики испытаний и обеспечивает информационное взаимодействие имитаторов бортовых систем с оборудованием командно-измерительной системы. Диаграмма формирования методики испытаний командно-программного управления бортовой аппаратурой показана на рисунке 2.5.

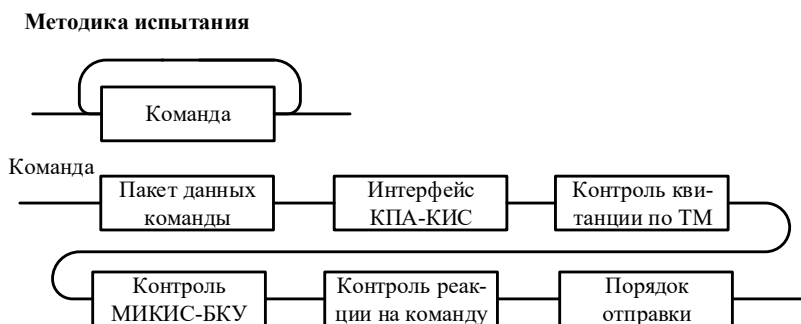


Рисунок 2.5 - Диаграмма формирования методики испытаний командно-программного управления бортовой аппаратурой космического аппарата

Формирование методики испытаний сводится к заданию настроек блоков «Пакет данных команды», «Интерфейс КПА-КИС» и «Порядок отправки», которые определяют поведение имитаторов в соответствии с выбранным вариантом командно-программного управления. Критерии анализа функционирования бортовой аппаратуры задаются в контрольных точках: «Контроль квитанции по ТМ», «Контроль МИ КИС БКУ», «Контроль реакции на команду». Такой подход, в отличие от метода формирования сценариев, позволяет легко и быстро создавать испытания и выполнять анализ прохождения команд.

Разработанный метод реализован в подсистеме испытаний командно-программного управления бортовой аппаратурой космического аппарата, входящей в состав программного обеспечения контрольно-проверочной аппаратуры (рисунок 2.6).

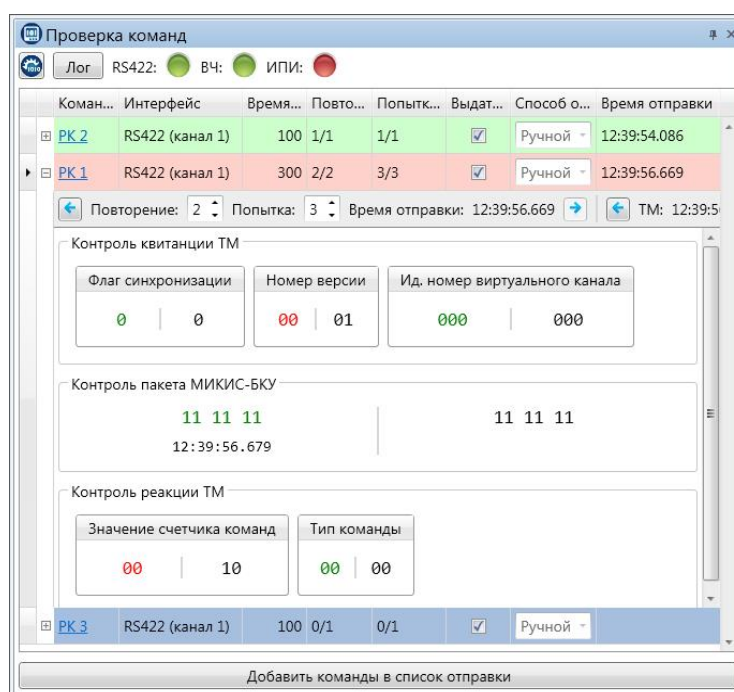


Рисунок 2.6 - Подсистема испытаний командно-программного управления

Программное обеспечение выполняет заданные в методике испытаний последовательности команд и отображает результаты анализа контрольных точек. Ведется журнал выполняемых действий, передаваемых и получаемых данных с возможностью формирования отчета по всем проводимым испытаниям.

Имитационное моделирование командно-программного управления бортовой аппаратурой позволило разработать новый метод автоматизации испытаний оборудования командно-измерительной системы космического аппарата. При проведении имитационных экспериментов выполнено исследование особенностей функционирования

бортовых систем, разработаны структуры данных для задания команд и телеметрии, сформирован перечень команд и критерии контроля их прохождения, созданы алгоритмы информационного взаимодействия имитаторов бортовых систем. Созданные алгоритмы и методы реализованы в программном обеспечении подсистемы испытаний командно-программного управления бортовой аппаратурой космического аппарата.

### *3 Разработка методов оценки различных аспектов кибербезопасности информационно-телекоммуникационных систем*

Проведены исследования по разработке методов оценки различных аспектов кибербезопасности информационно-телекоммуникационных систем. При использовании интернет-ресурсов накапливается большое количество данных, как на стороне пользователей, так и на стороне интернет-провайдеров. Обработка таких данных с привлечением различных методов Big-data, позволяет получить новые знания о поведении интернет-пользователей. Выявление аномалий потребления ресурсов основано на анализе результатов кластеризации пользователей Интернет-ресурсов за разные промежутки времени и может свидетельствовать об активности компьютерных вирусов, утечках информации и прочих угрозах кибербезопасности организаций.

В качестве исходных данных использованы журналы сервера доступа в интернет за месяц. Общий объем данных составляет около 8 Гигабайт, а количество обрабатываемых запросов около 100 миллионов. Полученные данные были нормализованы с помощью введения справочников доменов и пользователей. Затем выполнено понижение размерности пространства характеристик, с помощью перехода от адресов ресурсов (около 4000) к тематикам сайтов (около 160). Полученные данные были разбиты на кластеры, с помощью метода иерархической кластеризации. При расчете расстояний использовался алгоритм минимизации дисперсии Уорда. Наиболее показательными оказались разбиения на 2 и 3 кластера. Для визуализации полученных результатов использованы дендрограммы, метод главных компонент (PCA) и линейный дискриминантный анализ (LDA). Наиболее наглядный результат для трех и большего количества кластеров показал метод LDA (рисунок 3.1).

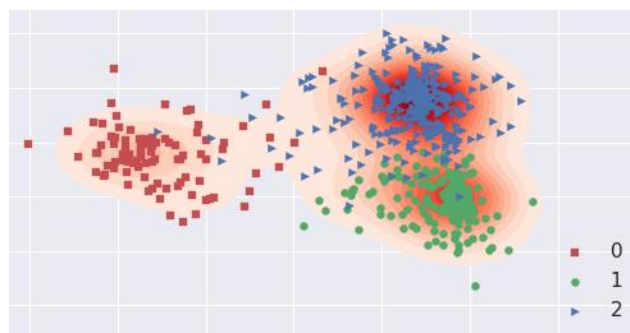


Рисунок 3.1 - Результат применения LDA для визуализации 3-х кластерной структуры пользователей Интернет-ресурсов

Предложенная методика анализа может быть применена для обнаружения аномалий использования интернет ресурсов, как визуально, так и с помощью выделения одиночных наиболее удаленных элементов. Обнаруженные аномалии могут рассматриваться как один из аспектов снижения кибербезопасности. Дополнительно, может быть реализован динамический контроль распределения пользователей по кластерам. Признаком опасного изменения является переход пользователя в отдельный кластер. Перспективными представляются дальнейшие исследования в данном направлении.

Проведены работы по мониторингу кибербезопасности корпоративной сети Федерального исследовательского центра «Красноярский научный центр СО РАН» и развитию моделей безопасности интернет-сервисов. Апробированы и применены новые методы противодействия массовым рассылкам, основанные на анализе корреспонденции публичных адресов, за счет которых удалось в несколько раз уменьшить процент спама (рисунок 3.2.).

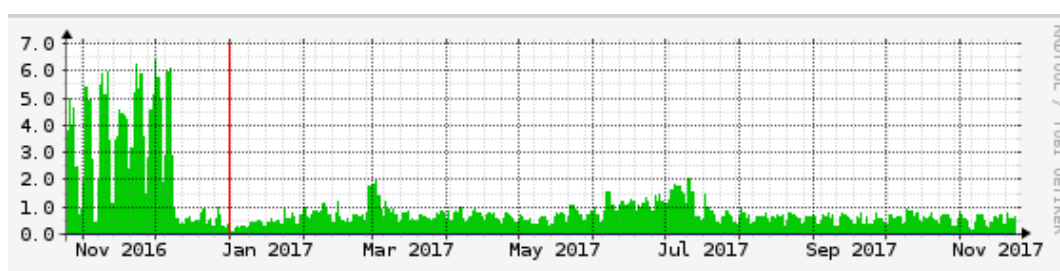


Рисунок 3.2 - Скорость поступления писем, идентифицированных как спам

С целью повышения эффективности функционирования систем информационной безопасности в настоящее время активно развиваются системы управления событиями информационной безопасности (Security information and event management). Основная их



функция – координация действий и отражение актуального состояния всех систем, что полностью согласуется с концепцией кибербезопасности. Разработана функциональная схема универсальной сигнатурной SIEM системы (рисунок 3.3), которая может настраиваться с учетом технических, программных и правовых особенностей организации. Конфигурирование производится наполнением СУБД, содержащей общие параметры системы, сигнатуры угроз и сценарии реагирования на выявленные угрозы, а также может содержать базу знаний, содержащую прецеденты угроз для дальнейшего предотвращения. Переход к использованию системы управления событиями информационной безопасности позволяет в полной мере осуществлять контроль за состоянием различных систем безопасности, которые в комплексе позволяют наиболее качественно выявлять и противодействовать различным киберугрозам организации.

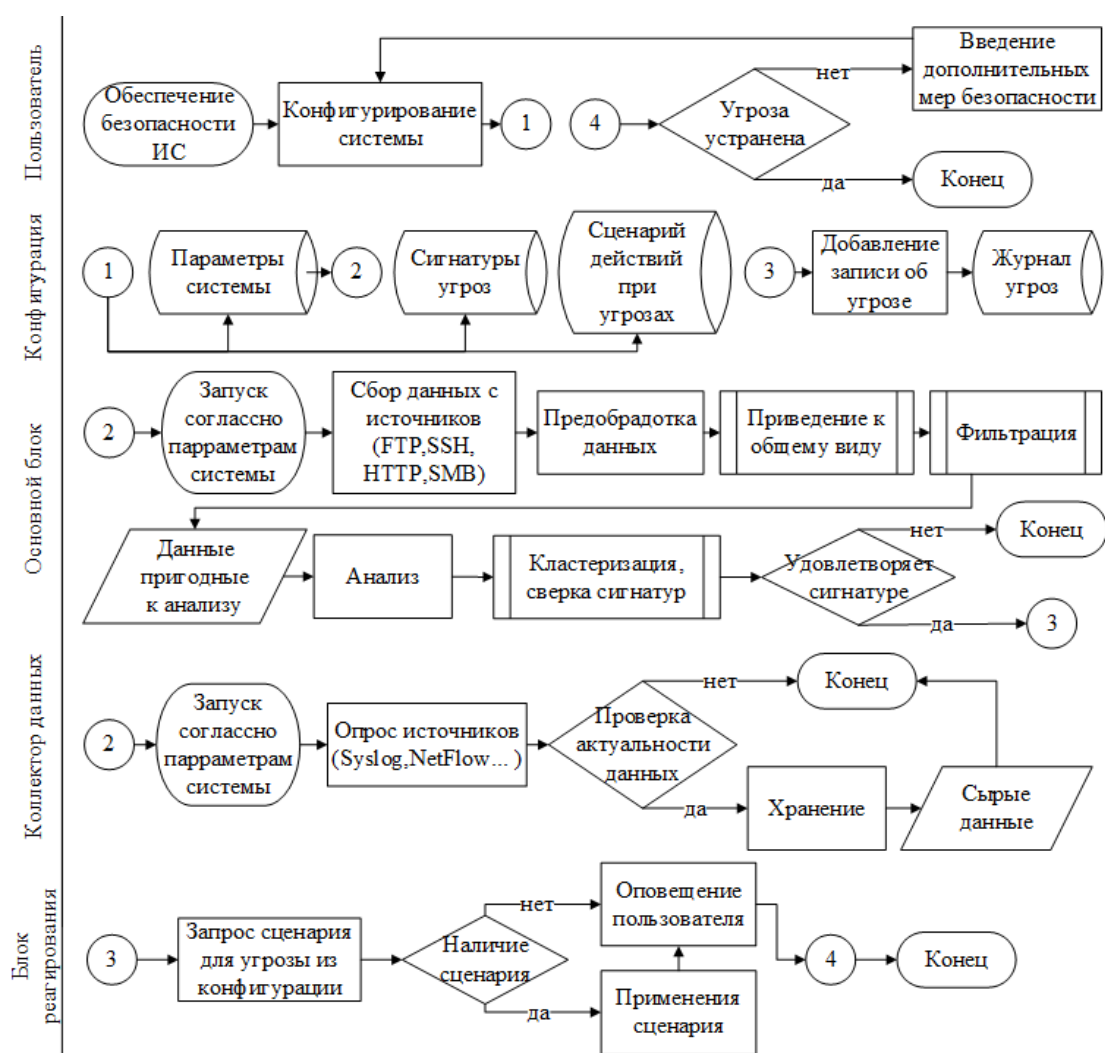


Рисунок 3.3 - Модель универсальной SIEM системы

В рамках развития методики разработки защищенных веб-приложений предложен метод интегральной оценки защищенности веб-приложений. Метод основывается на

международном стандарте ASVS (Advanced Security Verification Standard), который содержит требования безопасности, применяемые при проектировании, разработке и тестировании современных веб-приложений. Данный стандарт определяет три уровня верификации в зависимости от типа системы. Каждый уровень верификации определяет набор критериев, которые необходимо проверить для оценки защищенности системы. Критерии оценки стандарта ASVS сгруппированы по категориям, например, аутентификация, управление сессиями, безопасность HTTP и т.д. В рамках развития метода оценки защищенности веб-приложений в данный набор включена группа критериев для проверки модели безопасности, которая включает следующие критерии: 1) проверки выбранной модели безопасности на соответствие целям и задачам; 2) проверки элементов модели на полноту охвата предметной области; 3) проверки элементов модели на непротиворечивость; 4) проверки алгоритмов контроля доступа на покрытие тестовыми данными. Разработан метод интегральной оценки защищенности веб-приложений на основе векторов групповых оценок критериев. Для подсчета групповых оценок используется квадрат Евклидова расстояния между нулевым вектором  $0$  и вектором из нормализованных значений критериев  $K_i = (k_{i1}, k_{i2}, \dots, k_{im_i})$  в интервале  $[0,1]$ :  $g_i = \sqrt{\sum_{j=1}^{m_i} k_{ij}^2} / m_i$ . Чем ближе полученное значение  $g_i$  к 1, тем больше критериев внутри группы удовлетворяют заданным требованиям. На основе вектора групповых оценок безопасности построена интегральная оценка безопасности с использованием средневзвешенного значения:  $I = \sum_{i=1}^n \omega_i g_i$ , где  $\sum_{i=1}^n \omega_i = 1$ . Коэффициенты  $\omega_i$  назначаются экспертами для каждой группы критериев, что позволяет увеличить или уменьшить вес определенной группы в итоговой оценке. Данный подход обеспечивает гибкость при оценке информационных систем, в которых компоненты вносят разные веса в итоговую оценку. Разработанный метод оценки безопасности позволяет повысить уровень защищенности современных веб-приложений.

#### *4 Разработка статистических методов распознавания образов и восстановления многомерных зависимостей в условиях неоднородных данных, основанных на принципах коллективного оценивания и гибридного моделирования*

Разработана методика синтеза и анализа регрессионной оценки многомерной решающей функции для двувальтернативной задачи распознавания образов в условиях обучающих выборок большого объёма и неоднородных данных. Её синтез осуществляется путём декомпозиции исходной информации и анализа количественных характеристик получаемых множеств случайных величин на основе кривой регрессии. Определены условия асимптотической несмещённости и состоятельности регрессионной оценки многомерной решающей функции относительно оптимального байесовского уравнения разделяющей поверхности между классами. На этой основе определена аналитическая зависимость показателей качества аппроксимации исследуемой решающей функции от методов дискретизации пространства признаков классифицируемых объектов и их размерности, что позволило сформировать эффективную методику оптимизации непараметрических алгоритмов распознавания образов.

Исследован альтернативный подход построения непараметрических уравнений разделяющих поверхностей между классами в условиях больших выборок, основанный на декомпозиции исходных данных по их объёму, использовании принципов коллективного оценивания и технологии параллельных вычислений. Установлены условия асимптотической сходимости предложенного линейного коллектива решающих функций в двувальтернативной задаче распознавания образов и определена их зависимость от количества его составляющих. По сравнению с традиционной непараметрической решающей функции исследуемый коллектив имеет меньшую дисперсию, но несколько большее значение асимптотического выражения среднеквадратического отклонения. Это объясняется увеличением смещения коллектива непараметрических решающих функций, так как снижаются объёмы обучающих выборок при оценивании его составляющих. Неравномерность распределения элементов обучающей выборки между классами снижает аппроксимационные свойства коллектива непараметрических решающих функций. Данная тенденция особенно характерна для малых размерностей признаков классифицируемых объектов.

Методика синтеза гибридных систем распознавания образов, сочетающих преимущества параметрических и локальных аппроксимаций решающих функций, обобщена на условия неоднородных данных. Обучающая выборка в подобных условиях формируется при расширении возможности системы контроля, наличии заполненных

пропусков данных, нестационарности классифицируемых объектов и разнотипности их признаков. Идея предлагаемого подхода, например, при наличии заполненных пропусков данных состоит в выполнении следующих действий:

- осуществить декомпозицию неоднородной обучающей выборки  $V$  на множества их строк с заполненными пропусками данных  $V_1$  и без таковых  $V_2$ ;
- по выборке  $V_1$  построить непараметрическую оценку уравнения разделяющей поверхности  $\bar{f}_{12}(x)$  и соответствующий алгоритм распознавания образов  $\bar{m}_{12}(x)$  в двувальтернативной задаче классификации;
- с использованием выборки  $V_2$  и алгоритма  $\bar{m}_{12}(x)$  организовать вычислительный эксперимент и сформировать данные  $V'_2$ , включая значения признаков  $X$  классифицируемых объектов и корректирующей функции  $q(x)$ , характеризующей расхождение между «решениями»  $\bar{\sigma}(x)$  алгоритма  $\bar{m}_{12}(x)$  и «указаниями учителя»  $\sigma(x)$  (при наличии ошибки корректирующая функция принимает значение обратное по знаку уравнения разделяющей поверхности  $\bar{f}_{12}(x)$  и превышает его на величину  $\Delta$ );
- восстановить по выборке  $V'_2$  корректирующую функцию  $q(x)$  на основе непараметрической регрессии  $\tilde{q}(x)$ ;
- построить гибридный алгоритм распознавания образов  $\bar{\bar{m}}_{12}(x)$  с использованием решающей функции вида  $\bar{\bar{f}}_{12}(x) = \bar{f}_{12}(x) + \tilde{q}(x)$ .

Если принимаются несколько градаций по степени заполнения пропусков данных в обучающей выборке, то в гибридной системе распознавания образов формируется многоуровневая процедура корректировки уравнения разделяющей поверхности между классами.

Разработаны и исследованы модификации гибридных алгоритмов распознавания образов в условиях неоднородных данных, отличающиеся условиями классификации и видом корректирующих решающих функций.

Исследованы аппроксимационные свойства линейных коллективов многомерных непараметрических регрессий в условиях равномерного и неравномерного распределения исходных статистических данных между их составляющими. Установлены свойства их асимптотической несмещённости, сходимости. Полученные результаты позволили получить аналитические зависимости показателей качества аппроксимации коллектива непараметрических регрессий от объёма исходных данных, их размерности и количества составляющих коллектива. Неравномерное распределение исходных данных между составляющими коллектива по сравнению с равномерным распределением способствует

снижению значений среднеквадратического отклонения и увеличению дисперсии. По сравнению с традиционной непараметрической регрессией, их коллективы характеризуются меньшей дисперсией. Полученные результаты имеют большое значение при проектировании непараметрических моделей стохастических зависимостей в условиях больших массивов статистических данных с использованием технологии параллельных вычислений.

Аппроксимационные свойства непараметрической оценки плотности вероятности Розенблатта-Парзена  $\bar{p}(x)$  зависят от выбора коэффициентов размытости ядерных функций. Проведён анализ методов оптимизации  $\bar{p}(x)$  путём выбора значений коэффициентов размытости из условий минимума среднеквадратического отклонения непараметрической оценки плотности вероятности (метод 1) и максимума функции правдоподобия (метод 2). Эффективность сравниваемых методов оптимизации непараметрической оценки плотности вероятности при восстановлении равномерного закона распределения случайной величины сопоставима. Они характеризуются близкими значениями среднеквадратического отклонения оценки плотности вероятности. Выбор коэффициента размытости ядерной функции из условия минимума оценки среднеквадратического критерия позволяет получить меньшее значение дисперсии  $\bar{p}(x)$ . При этом смещение непараметрической оценки плотности вероятности значительно больше, если при её оптимизации использовать критерий максимума функции правдоподобия. В этих условиях более предпочтительно использовать метод 1 оптимизации непараметрической оценки плотности вероятности. При восстановлении нормального закона распределения в условиях больших объёмов статистических данных значения смещения, дисперсии и среднеквадратического отклонения  $\bar{p}(x)$  с использованием сравниваемых методов оптимизации отличаются несущественно. При  $n < 200$  и ступенчатой ядерной функции целесообразно использовать метод 1, который обеспечивает значительно меньшее значение дисперсии  $\bar{p}(x)$ . Если основу синтеза  $\bar{p}(x)$  составляет ядерная функция Епанечникова, то в условиях малых выборок более предпочтительным является метод 2. С уменьшением области определения плотности вероятности смещение, дисперсия и среднеквадратическое отклонение  $\bar{p}(x)$  увеличиваются, но их соотношения для сравниваемых методов оптимизации и приведённые выше рекомендации сохраняются. Полученные результаты открывают возможность сравнения традиционных методов с рандомизированным подходом оптимизации непараметрической оценки плотности вероятности. Их отличие состоит в признании случайной природы коэффициентов размытости ядерных функций. Тогда

задача оптимизации непараметрической оценки плотности вероятности может быть сведена к выбору оптимальных параметров закона распределения коэффициентов размытости ядерных функций.

Проведено сравнение оптимального и эвристических методов дискретизации области значений двумерной случайной величины. Формула оптимальной дискретизации получена ранее из условия минимума среднеквадратического отклонения регрессионной оценки плотности вероятности. Эвристические рекомендации  $D(m)$  основаны на введении минимального количества  $m$  наблюдений случайной величины, принадлежащих двумерному интервалу дискретизации области её значений. Для нормального закона распределения зависимых случайных величин количество оптимальных интервалов дискретизации области их значений зависит от коэффициента корреляции, объёма исходных данных и метода дискретизации. С ростом абсолютного значения коэффициента корреляции  $r$  количество оптимальных интервалов дискретизации увеличивается. При малых значениях  $|r| < 0,5$  эффективность оптимальной формулы дискретизации сопоставима с методом  $D(m=1)$ . В условиях  $|r| > 0,5$  целесообразно использовать методы дискретизации  $D(m=3)$  и  $D(m=5)$ . Отмеченная закономерность не зависит от значений математических ожиданий и среднеквадратических отклонений двумерных случайных величин. Предлагаемая методика может быть использована при сравнении эффективности методов дискретизации области значений двумерной случайной величины с другими законами распределения. Полученные результаты имеют важное значение при восстановлении плотности вероятности, их доверительного оценивания и проверке гипотез о распределениях случайных величин с использованием критерия Пирсона.

Программы «Восстановление стохастической зависимости в условиях неопределенности», «Программа для реализации непараметрических алгоритмов распознавания образов парзеновского типа в многомерном пространстве признаков классифицируемых объектов» получили свидетельства о государственной регистрации программ для ЭВМ. Их эффективность подтверждена успешным применением при оценивании состояния лесных массивов по данным дистанционного зондирования.

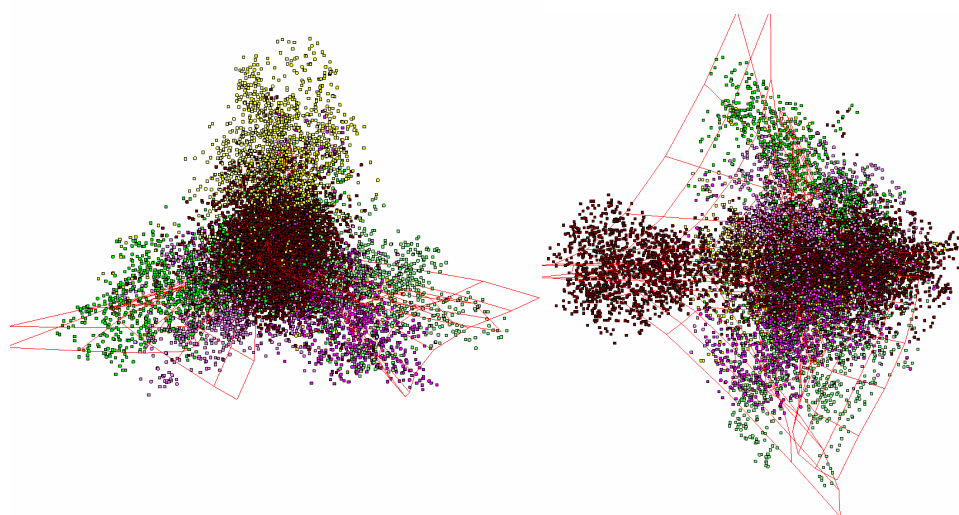
*5 Выявление новых структурных элементов в символьных последовательностях, соответствующих генетическим данным, на основе изучения взаимного распределения триплетов и октаплетов*

Рассмотрены генетические последовательности длины  $L$ , состоящие из символов алфавита  $M = \{A, C, G, T\}$ . Для последовательности символов составлен частотный словарь толщины 3, содержащий список всех триплетов  $v_1 v_2 v_3$  идущих подряд символов с указанием частот этих триплетов. Число таких комбинаций 64. В качестве частоты рассматривается отношение количества копий  $n_\omega$  выбранного триплета к общему количеству всех триплетов  $N$ , где  $N$  – сумма всех  $n_\omega$ :  $f_\omega = \frac{n_\omega}{N}$ . Любой частотный словарь  $W_3$  ставит в соответствие геному множество точек в 64-мерном метрическом пространстве.

Для обнаружения структуры в генетической последовательности проведена предварительная обработка, позволяющая поставить в соответствие данной последовательности множество точек в 64-мерном пространстве триплетов. Последовательность просканирована окном длины  $\Delta$  с шагом  $t$ . Для каждого положения  $i$  рамки определен участок генетической последовательности, совпадающий с рамкой считывания, для которого вычислен частотный словарь  $W_3^{(i)}$ , соотносящийся с  $i$ -ой точкой 64-мерного пространства. С каждой точкой 64-мерного пространства связаны следующие параметры: номер центрального символа рассматриваемого участка и относительная фаза. Номер центрального символа участка совпадает с номером этого символа в последовательности. Относительная фаза определяется тем, попал рассматриваемый участок в кодирующую или не кодирующую область последовательности. Участок относится к кодирующим, если он целиком попадал в кодирующую область последовательности. Для не кодирующего участка соответствующая ему точка помечается символом  $J$ . Если участок относится к кодирующим, для него возможны 6 вариантов маркировки:  $B_0, B_1, B_2, F_0, F_1, F_2$ . Для кодирующего участка, аннотированного в последовательности как считывающийся в прямом направлении, вычислен остаток от деления на 3 разности номеров центрального символа участка и первого символа кодирующей области, к которой он относится. В соответствии с величиной остатка от деления точка помечена символом  $B_0, B_1$  или  $B_2$ . Для участка, аннотированного как считывающийся в обратном направлении, вычисляется остаток от деления на 3 разности номеров последнего символа кодирующей области, к которой относится участок, и центрального символа участка. В зависимости от значения остатка от деления точка

помечалась символами  $F_0, F_1$  или  $F_2$ . Для всех генетических последовательностей длина рамки считывания  $\Delta = 603$ , шаг  $t = 11$ . Для каждого генома в пространстве частот при помощи программы VidaExpert построены проекции из 64-мерного пространства частот троек в пространство первых трёх главных компонент, построенных по этому множеству точек. Рассматривались проекции на плоскость пространства первых двух главных компонент и второй и третьей главной компоненты. Чтобы отобразить принадлежность точек к не кодирующим фрагментам и выделить относительные фазы, точки помечены разными цветами: точкам, относящимся к не кодирующим участкам, соответствует коричневый цвет; точкам, относящимся к фрагментам с относительными фазами  $B_0$  и  $F_0$ , соответствуют темно-малиновый и светло-малиновым цвет; фрагментам с относительными фазами  $B_1$  и  $F_1$  соответствуют темно-зеленый и светло-зеленый цвет; фрагментам с относительными фазами  $B_2$  и  $F_2$  соответствуют темно-желтый и светло-желтый цвет.

В результате исследования роанализировано 185 геномов хлоропластов наземных растений, 76 геномов одноклеточных и многоклеточных водорослей и 7 геномов цианобактерий, размещенных в EMBL-банке. Обнаружено, что все геномы хлоропластов наземных растений в плоскости первой и второй главных компонент имеют выраженную трех лучевую структуру. Для подавляющего большинства геномов типичен вид в плоскости первых двух главных компонент и второй и третьей главной компоненты, который показан на рисунке 5.1.



а) Проекция в плоскость первых  
двух главных компонент

б) Проекция в плоскость второй  
и третьей главных компонент

Рисунок 5.1 - Типичный вид структуры данных геномов наземных растений  
в проекциях пространства первых трех главных компонент

(приведена структура генома *Nimphaea alba*)

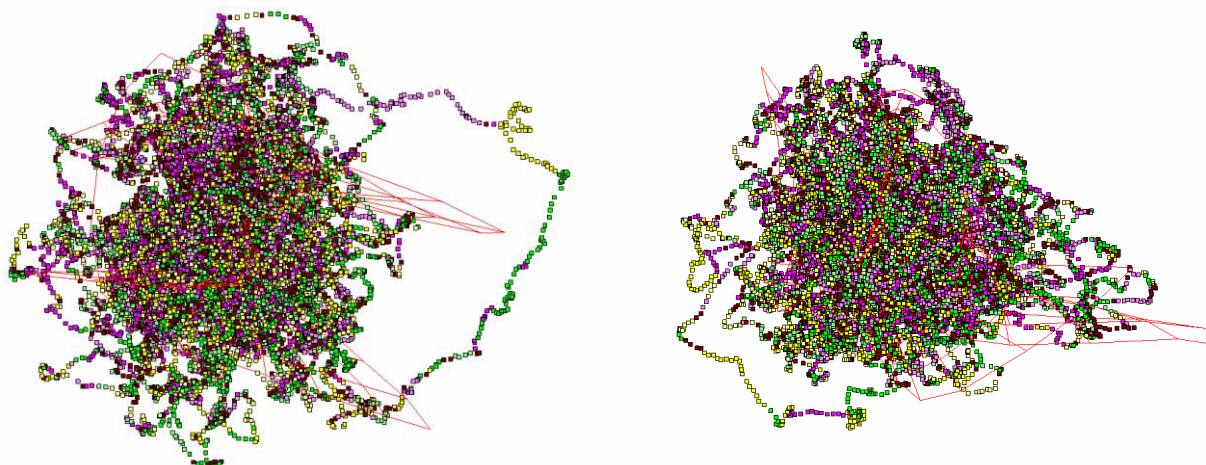


Из рисунка 5.1а видно, что структура генома хорошо кластеризована относительно точек из кодирующих и не кодирующих участков генома. Центральную часть занимают точки из некодирующих участков, они окрашены в коричневый цвет. Один луч состоит из точек, относящихся к прямой и обратной относительной фазе 2 (точки, отмеченные светло желтым и темно желтым цветом), второй луч содержит точки, относящиеся к прямой относительной фазе 0 и обратной относительной фазе 1 (точки темно сиреневого и светло зеленого цвета), и третий состоит из точек, относящихся к обратной относительной фазе 0 и прямой относительной фазе 1 (точки светло сиреневого и темно зеленого цвета). Из рисунка 5.1б видно, что точки, занимающие центральную часть распределения в проекции рисунка 5.1а, разделены на два кластера. Первый находится в центре трёхлучевой структуры, а второй обособлен от остальных точек. Такой типичной структурой данных в пространстве первых трёх главных компонент обладают геномы 179 хлоропластов из 185 рассматриваемых. Наличие разбиения фрагментов на два кластера позволяет утверждать существование восьмикластерной, а не семикластерной (Gorban et al., 2003, 2005) структуры. У оставшихся геномов хлоропластов наблюдались следующие отличия от типичной структуры данных. У геномов хлоропластов *Selaginella moellendorffii* и *Selaginella uncinata* отсутствует обособленная группа точек, относящаяся к некодирующим участкам генома. Это представители древних споровых растений, которые впервые появились около 400 млн. лет назад. Структура данных этих геномов существенно больше похожа на структуру данных геномов бактерий. У растений *Brachypodium distachyon*, *Juniperus bermudiana*, *Pinus contorta*, *Jacobaea vulgaris*, *Najas flexilis* относительные фазы в лучах перемешаны.

Для большинства геномов хлоропластов одноклеточных и многоклеточных водорослей установлено, что структура данных в пространстве первых трех главных компонент очень похожа: она имеет центральное ядро из фрагментов некодирующих участков и три луча, состоящих из фрагментов кодирующих участков. Кроме того, выделяется отдельная группа фрагментов, не входящая в трехлучевую структуру. Обнаружены следующие отклонения от типичной структуры: трехлучевая структура и отсутствие обособленного кластера, шестилучевая структура и отсутствие обособленного кластера, шестилучевая структура и наличие обособленного кластера. Большой разброс по структуре геномов хлоропластов для водорослей по сравнению с наземными растениями может объясняться гораздо большими различиями в среде обитания у водорослей.

Гипотеза о происхождении хлоропластов от одноклеточных свободноживущих фотосинтезирующих бактерий позволяет ожидать, что структура геномов цианобактерий будет подобна аналогичной структуре хлоропластов. Рассмотрены структуры для 7

геномов цианобактерий. На рисунке 5.2 показан характерный вид структуры геномов цианобактерий в проекции на плоскость первых двух главных компонент. Как видно из рисунка, структура геномов цианобактерий представляет собой своеобразные клубки из цепочек, точки в которых расположены в той же последовательности, в которой соответствующие участки расположены в геноме.



а)

б)

Рисунок 5.2 - Структура данных *Microcystis aeruginosa* (а) и *Nostoc sp. PCC 7107* (б)  
в проекции на плоскость первых двух главных компонент

Таким образом, исследование показало, что структура геномов цианобактерий существенно отличается от структуры геномов хлоропластов. Это наблюдение может свидетельствовать о том, что расхождение от одноклеточных свободноживущих фотосинтезирующих бактерий современных цианобактерий и хлоропластов произошло очень давно, и о независимом характере их развития.

## ЗАКЛЮЧЕНИЕ

За отчетный период выполнения проекта получены следующие результаты.

Выполнены работы по созданию новых и развитию существующих методов и алгоритмов аналитической обработки и интеграции разнородных данных. Выполнен анализ вычислительных затрат расчета аналитических показателей при реализации многошаговых методик. Предложена метамодель формирования многомерных оценок показателей природно-техногенной безопасности на основе спецификации Common Warehouse Metamodel (CWM). Выполнено развитие методов и средств оперативной аналитической обработки пространственных данных. Для консолидации разнородных данных в структуру централизованного хранилища разработан подход к представлению древовидной структуры XSD файла в виде объектной модели на основе расширения стандарта CWM, выполнена программная реализация предложенного подхода. Предложен технологический подход к организации межсистемного информационного взаимодействия между гетерогенными системами, основанный на унифицированном представлении разнородных структур обмена данными, разработана архитектура программных средств организации межсистемного взаимодействия.

Выполнено развитие web-технологий открытых архивов на примере тестовой модели институционального репозитория ИВМ СО РАН. Институциональные репозитории, формируемые в технологиях архивов открытого доступа, позволят произвести качественный учет и продвижение научных публикаций, обеспечить открытость результатов научных исследований для мирового сообщества, повысить индивидуальные показатели цитирования. Открытый архив позволит в перспективе успешно развивать связывание данных, а также интегрировать данные репозитория в корпоративные, региональные или национальные проекты открытых архивов, в соответствие с общемировыми тенденциями развития web-технологий.

Выполнено развитие оригинальной реализации технологии модельно-ориентированной разработки программного обеспечения (Model Driven Development, MDD), предложена модель взаимодействия клиентской и серверной частей программного комплекса модельно-ориентированной системы, разработан алгоритм формирования пользовательского интерфейса на основе метаданных описывающих управляющую модель. Использование алгоритма позволяет достичь высокого уровня адаптируемости информационной системы.

Разработан метод автоматизации испытаний на основе имитационной модели командно-программного управления бортовой аппаратуры космического аппарата. Разработаны алгоритмы информационного взаимодействия имитаторов бортовых систем и алгоритм анализа прохождения команд по контролю телеметрии. На основе построенной имитационной модели разработан метод автоматизации испытаний командно-программного управления бортовой аппаратурой космического аппарата. Разработанный метод реализован в подсистеме испытаний командно-программного управления бортовой аппаратурой космического аппарата, входящей в состав программного обеспечения контрольно-проверочной аппаратуры. Имитационное моделирование командно-программного управления бортовой аппаратурой позволило разработать новый метод автоматизации испытаний оборудования командно-измерительной системы космического аппарата. При проведении имитационных экспериментов выполнено исследование особенностей функционирования бортовых систем, разработаны структуры данных для задания команд и телеметрии, сформирован перечень команд и критерии контроля их прохождения, созданы алгоритмы информационного взаимодействия имитаторов бортовых систем. Созданные алгоритмы и методы реализованы в программном обеспечении подсистемы испытаний командно-программного управления бортовой аппаратурой космического аппарата. Разработаны графические интерфейсы программной инфраструктуры имитационного моделирования бортовой аппаратуры спутниковых систем.

Разработаны методы и принципы оценки аспектов кибербезопасности в информационно-телекоммуникационных системах. Предложена методика обработки и анализа больших объемов данных по использованию интернет-ресурсов и выявлению аномалий их использования для снижения уровня киберугроз. Разработана функциональная схема универсальной сигнатурной системы управления событиями информационной безопасности, позволяющей в полной мере осуществлять контроль за состоянием различных систем безопасности путем интеграции в единую систему всех источников данных и на основе совместного их анализа. Предложен метод интегральной оценки защищенности веб-приложений на основе векторов групповых оценок критериев, обеспечивающий гибкость при оценке информационных систем, в которых компоненты вносят разные веса в итоговую оценку. Применение перечисленных методов позволяет проводить комплексную оценку кибербезопасности организации и, применяя адекватные методы защиты, эффективно выявлять и противодействовать киберугрозам на уровне информационно-телекоммуникационным систем.

Разработаны статистические методы распознавания образов и восстановления многомерных зависимостей в условиях больших выборок неоднородных данных, основанные на дискретизации области их определения, принципах коллективного оценивания и гибридного моделирования. Предложенные непараметрические и гибридные решающие функции обладают свойствами асимптотической сходимости. Установлена их аналитическая зависимость от особенностей исходных данных и методов их обработки. Полученные выводы обобщены при синтезе линейных коллективов непараметрических регрессий. На этой основе разработана методика синтеза эффективной структуры непараметрических и гибридных систем распознавания образов, восстановления стохастических зависимостей в условиях больших выборок неоднородных данных. Решены проблемы дискретизации области значений двумерной случайной величины и оптимизации непараметрической оценки плотности вероятности, обеспечивающих вычислительную эффективность коллективов решающих функций и гибридных моделей. Разработаны программные средства синтеза непараметрических алгоритмов распознавания образов и восстановления многомерных стохастических зависимостей, которые используются в задаче оценивания состояния лесных массивов по данным дистанционного зондирования.

Исследована внутренняя структурированность геномов хлоропластов и цианобактерий на основе кластеризации частотных словарей триплетов отдельных фрагментов генома, определяемых регулярным порядком вне зависимости от функциональной роли участка с помощью метода упругих карт. Обнаружено, что подавляющее большинство геномов хлоропластов имеет восьмикластерную структуру, геномы цианобактерий имеют одинаковую структуру, представляющую собой своеобразный клубок из нитей, образованных точками, соответствующими последовательным фрагментам генома. Выявленная структурированность существенно отличается от ранее обнаруженной для бактериальных геномов.

План работ по проекту выполнен полностью на высоком научном уровне, что подтверждается публикациями коллектива исполнителей проекта в ведущих отечественных и зарубежных научных журналах.

## ПРИЛОЖЕНИЕ А

Научные публикации в журналах, индексируемых в российских и международных информационно-аналитических системах научного цитирования

1. Erunova M., Sadovsky M. GIS-aided modelling of two Siberian reservation sites // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – V. 10208 LNCS. – P. 617-628. DOI: 10.1007/978-3-319-56148-6\_55.
2. Kireeva O., Bushmelev E., Emelianchik E., Salmina A., Sadovsky M. An impact of severe preeclampsia on cardiovascular system adaptation of newborns in early neonatal period // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – V. 10208 LNCS. – P. 517-523. DOI: 10.1007/978-3-319-56148-6\_46.
3. Sadovsky M., Fontaine J.-F., Andrade-Navarro M.A., Yakubailik Y., Rudenko N. Lost strings in genomes: What sense do they make? // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – V. 10209 LNCS. – P. 20-29. – DOI: 10.1007/978-3-319-56154-7\_3.
4. Penkova T., Kochetkov S. Unified Hierarchical Model of the Data Exchange // Advances in Intelligent Systems Research (AMMS2017) / Atlantis Press. – 2017. – V. 153. – P. 143-146. doi: 10.2991/amms-17.2017.33
5. Ковязина Е.В. Открытый архив в научной библиотеке // Научные и технические библиотеки, 2017. - № 2. – С.57-64. DOI: wos:000405549000007
6. Kononov D., Isaev S. Development of extended path-based role access control model for web applications // CEUR Workshop Proceedings, 2017 – Vol.1839. – P.166-173. DOI: 2-s2.0-85020551584
7. Penkova T., Nicheporchuk V., Metus A. Comprehensive operational control of the natural and anthropogenic territory safety based on analytical indicators // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – V. 10313 LNAI. – P. 263-270. DOI: 10.1007/978-3-319-60837-2\_22
8. Lapko A.V., Lapko V.A. Comparison of the Efficiency of Methods of Digitizing the Range of Values of Dependent Random Variables During Synthesis of a Nonparametric Assessment of Two-Dimensional Probability Density // Measurement Techniques. – V. 60 (4). – P. 325-330. DOI: 10.1007/s11018-017-1196-1
9. Lapko A.V., Lapko V.A., Yuronen E.A. Effectiveness comparison of sampling methods of a range of values of a two-dimensional random value at probability density estimation // 2017

- International Siberian Conference on Control and Communications, SIBCON 2017 – Proceedings. – Art. 7998533. DOI: 10.1109/SIBCON.2017.7998533
10. Lapko A.V., Lapko V.A. Analysis of Optimization Methods for Nonparametric Estimation of the Probability Density with Respect to the Blur Factor of Kernel Functions // Measurement Techniques. – V. 60 (6). – P. 515-522. – DOI: 10.1007/s11018-017-1228-x
  11. Penkova T.G. Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety // Procedia Computer Science. – V. 112. – P. 99-108. – DOI: 10.1016/j.procs.2017.08.179.
  12. T. Penkova Modelling the Information Structure of the Document Based on Specialised Markup Language // Proceedings of the 2017 international conference on applied mathematics : Atlantis press, 2017. – Vol. 141: International Conference on Applied Mathematics, Modelling and (MAY 21-22, 2017, Beijing, PEOPLES R CHINA). – P. 66-69. DOI: wos:000416087700014
  13. L. Nozhenkova, O. Isaeva, E. Gruzenko Markov A. Architecture of the Simulation Modeling Infrastructure Based on the Simulation Model Portability Standard // Proceedings of the 2017 international conference on applied mathematics : Atlantis press, 2017. – Vol. 141: International Conference on Applied Mathematics, Modelling and (MAY 21-22, 2017, Beijing, PEOPLES R CHINA). – P. 125-128. DOI: wos:000416087700027
  14. L. Nozhenkova, O. Isaeva, E. Gruzenko Markov A. Problem-Oriented Integration of SMP-models in the Simulation Modeling Infrastructure // Proceedings of the 2017 international conference on applied mathematics : Atlantis press, 2017. – Vol. 141: International Conference on Applied Mathematics, Modelling and (MAY 21-22, 2017, Beijing, PEOPLES R CHINA). – P. 121-124. DOI: wos:000416087700026
  15. Korobko A., Nozhenkova L. Ordered multidimensional model construction of relational source for integral OLAP-modeling // Application of Information and Communication Technologies, AICT 2016 – Conference Proceedings – Art. 7991670. DOI: 10.1109/ICAICT.2016.7991670
  16. Ноженкова Л.Ф., Исаева О.С., Вогоровский Р.В., Мишур А.В. Формирование процедур внешнего командно программного управления для испытаний командно-измерительной системы космического аппарата // Исследования наукограда. – Железногорск: Ассоциация "Технологическая платформа "Национальная информационная спутниковая система", 2017. – Т. 1, № 4. – С. 175-183. DOI: 10.26732/2225-9449-2017-4-175-183
  17. Nozhenkova L., Isaeva O., Markov A., Koldyrev A., Vogorovski R., Evsyukov A. Simulation infrastructure design on the basis of the space industry's international standards //

Advances in Intelligent Systems Research (CAAI2017) Atlantis Press. – 2017. – V. 134. – P. 138-141. doi: 10.2991/caai-17.2017.28

18. Nicheporchuk V.V., Chernyakova N.A. Application of data infrastructure for estimation of emergency risks // CEUR Workshop Proceedings, 2017. – Vol. 2033. – P.280-284. DOI: 2-s2.0-85040221656



## ПРИЛОЖЕНИЕ В

Выписка из плана научно-исследовательской работы на 2017 год

| Содержание работы                                                                                                                                                                                                                                                                                                                                                                   | Планируемый результат выполнения работы                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Развитие и апробация технологий создания интегрированных систем в различных областях территориального и корпоративного управления. Развитие методов автоматизированной поддержки испытаний программно-аппаратных комплексов бортовой аппаратуры космического аппарата. Разработка методов оценки различных аспектов кибербезопасности информационно-телекоммуникационных систем. | 1. Алгоритмы интеграции разнородных данных в структуру централизованного хранилища. Алгоритмы преобразования элементов управляющей модели в объекты программной реализации. Методы имитационного моделирования командно-программного управления функционированием командно-измерительной системы космического аппарата. Методы оценки аспектов кибербезопасности в информационно-телекоммуникационных системах. |
| 2. Разработка статистических методов распознавания образов и восстановления многомерных зависимостей в условиях неоднородных данных, основанных на принципах коллективного оценивания и гибридного моделирования                                                                                                                                                                    | 2. Статистические методы распознавания образов и восстановления многомерных зависимостей, основанные на принципах коллективного оценивания и гибридного моделирования.                                                                                                                                                                                                                                          |
| 3. Выявление новых структурных элементов в символьных последовательностях, соответствующих генетическим данным, на основе изучения взаимного распределения триплетов и октаплетов.                                                                                                                                                                                                  | 3. Список новых структурных элементов в символьных последовательностях, их комбинаторные и статистические свойства, связь между структурами и семантикой генома.                                                                                                                                                                                                                                                |